Seminal papers in nonlinear optimization

Nick Gould, CSED, RAL, Chilton, OX11 0QX, England (n.gould@rl.ac.uk)

December 7, 2006

The following papers are classics in the field. Although many of them cover topics outside the material we shall have time to cover, they are all worth reading.

Early quasi-Newton methods

These methods were introduced by

> W. Davidon, "Variable metric method for minimization", manuscript (1958), finally published *SIAM J. Optimization* **1** (1991) 1:17,

and championed by

> R. Fletcher and M. J. D. Powell, "A rapidly convergent descent method for minimization", *Computer J.* (1963) 163:168.

Although the so-called DFP method has been superseded by the more reliable BFGS method, it paved the way for a number of classes of important updates.

More modern quasi-Newton methods

Coincidentally, all of the papers

> C. G. Broyden, "The convergence of a class of double-rank minimization algorithms", *J. Inst. Math. Applcs.*, **6** (1970) 76:90,
>
> R. Fletcher, "A new approach to variable metric algorithms", *Computer J.* (1970) **13** (1970) 317:322,
>
> D. Goldfarb, "A family of variable metric methods derived by variational means", *Math. Computation* **24** (1970) 23:26, and
>
> D. F. Shanno, "Conditioning of quasi-Newton methods for function minimization", *Math. Computation* **24** (1970) 647:657

appeared in the same year. The aptly-named BFGS method has stood the test of time well, and is still regarded as possibly the best secant updating formula.

Quasi-Newton methods for large problems

Limited memory methods are secant-updating methods that discard old information so as to reduce the amount of storage required when solving large problems. The methods first appeared in

J. Nocedal, "Updating quasi-Newton matrices with limited storage", *Math. Computation* **35** (1980) 773:782, and

A. Buckley and A. Lenir, "QN-like variable storage conjugate gradients", *Math. Programming* **27** (1983) 155:175.

Secant updating formulae proved to be less useful for large-scale computation, but a successful generalization, applicable to what are known as partially separable functions, was pioneered by

A. Griewank and Ph. Toint, "Partitioned variable metric updates for large structured optimization problems", *Numerishe Mathematik* **39** (1982) 119:137, see also 429:448, as well as

A. Griewank and Ph. Toint, "On the unconstrained optimization of partially separable functions", in *Nonlinear Optimization 1981* (Powell, M., ed.) Academic Press (1982)

Conjugate gradient methods for large problems

Generalizations of Conjugate Gradient methods for non-quadratic minimization were originally proposed by

R. Fletcher and C. M. Reeves, "Function minimization by conjugate gradients", *Computer J.* (1964) 149:154, and

E. Polak and G. Ribiére, "Note sur la convergence de méthodes de directions conjuguées", *Revue Francaise d'informatique et de recherche opérationelle* **16** (1969) 35:43.

An alternative is to attempt to solve the (linear) Newton system by a conjugate-gradient like method. Suitable methods for terminating such a procedure while still maintaining fast convergence were proposed by

R. S. Dembo and T. Steihaug, "Truncated-Newton algorithms for large-scale unconstrained optimization", *Math. Programming* **26** (1983) 190:212.

Non-monotone methods

While it is usual to think of requiring that the objective function decreases at every iteration, this isn't actually necessary for convergence so long as there is some overall downward trend. The first method along these lines was by

L. Grippo, F. Lampariello and S. Lucidi, "A nonmonotone line search technique for Newton's method", *SIAM J. Num. Anal.*, **23** (1986) 707:716.

Trust-region methods

The earliest methods that might be regarded as trust-region methods are those by

K. Levenberg, "A method for the solution of certain problems in least squares", *Quarterly J. Appl. Maths*, **2** (1944) 164:168, and

D. Marquardt, "An algorithm for least-squares estimation of nonlinear parameters" *SIAM J. Appl. Maths*, **11** (1963) 431:441

for the solution of nonlinear least-squares problems, although they are motivated from the perspective of modifying indefinite Hessians rather than restricting the step. Probably the first "modern" interpretation is by

> S. Goldfeldt, R. Quandt and H. Trotter, "Maximization by quadratic hill-climbing", *Econometrica*, **34** (1966) 541:551.

Certainly, the earliest proofs of convergence are given by

> M. Powell, "A New Algorithm for Unconstrained Optimization", in *Nonlinear Programming*, (Rosen, J., Mangasarian, O., and Ritter, K., eds.) Academic Press (1970),

while a good modern introduction is by

> J. Moré, "Recent developments in algorithms and software for trust region methods", in *Mathematical Programming: The State of the Art*, (Bachem, A., Grötschel, M., and Korte, B., eds.) Springer Verlag (1983).

You might want to see our book

> A. Conn, N. Gould and Ph. Toint, "Trust-region methods", SIAM (2000)

for a comprehensive history and review of the large variety of articles on trust-region methods.

Trust-region subproblems

Almost all you need to know about solving small-scale trust-region subproblems is contained in the seminal paper

> J. Moré and D. Sorensen, "Computing a trust region step", *SIAM J. Sci. Stat. Comp.* **4** (1983) 533:572.

Likewise

> T. Steihaug, "The conjugate gradient method and trust regions in large scale optimization", *SIAM J. Num. Anal.* **20** (1983) 626:637

provides the basic truncated conjugate-gradient approach used so successfully for large-scale problems. More recently[1]

> N. Gould, S. Lucidi, M. Roma and Ph. Toint, "Solving the trust-region subproblem using the Lanczos method", *SIAM J. Optimization* **9** (1999) 504:525

show how to improve on Steihaug's approach by moving around the trust-region boundary. A particularly nice new paper by

> Y. Yuan, "On the truncated conjugate-gradient method", *Math. Programming*, **87** (2000) 561:573

proves that Steihaug's approximation gives at least 50% of the optimal function decrease when applied to convex problems.

The Symmetric Rank-One quasi-Newton approximation

Since trust-region methods allow non-convex models, perhaps the simplest of all Hessian approximation methods, the Symmetric Rank-One update, is back in fashion. Although it is unclear who first suggested the method,

[1] I would hate to claim "seminal" status for one of my own papers!

C. Broyden, "Quasi-Newton methods and their application to function minimization", *Math. Comp.* **21** (1967) 577:593

is the earliest reference that I know of. Its revival in fortune is due[2] to

A. Conn, N. Gould and Ph. Toint, "Convergence of quasi-Newton matrices generated by the Symmetric Rank One update" *Math. Programming*, **50** (1991) 177:196 (see also *Math. Comp.* **50** (1988) 399:430), and

R. Byrd, H. Khalfan and R. Schnabel "Analysis of a symmetric rank-one trust region method" *SIAM J. Optimization* **6** (1996) 1025:1039,

and it has now taken its place alongside the BFGS method as the pre-eminent updating formula.

Non-monotone methods

Non-monotone methods have also been proposed in the trust-region case. The basic reference here is the paper by

Ph. Toint, "A non-monotone trust-region algorithm for nonlinear optimization subject to convex constraints", *Math. Programming*, **77** (1997) 69:94.

Barrier function methods

Although they appear to have originated in a pair of unpublished University of Oslo technical reports by K. Frisch in the mid 1950s, (logarithmic) barrier function were popularized by

A. Fiacco and G. McCormick, "The sequential unconstrained minimization technique for nonlinear programming: a primal-dual method", *Management Science* **10** (1964) 360:366; see also *ibid* (1964) 601:617.

A full early history is given in the book

A. Fiacco and G. McCormick, "Nonlinear programming: sequential unconstrained minimization techniques" (1968), republished as *Classics in Applied Mathematics 4*, SIAM (1990).

The worsening conditioning of the Hessian was first highlighted by

F. Lootsma, "Hessian matrices of penalty functions for solving constrained optimization problems", *Philips Research Reports*, **24** (1969) 322:331, and

W. Murray, "Analytical expressions for eigenvalues and eigenvectors of the Hessian matrices of barrier and penalty functions", *J. Optimization Theory and Applics*, **7** (1971) 189:196,

although recent work by

M. Wright, "Ill-conditioning and computational error in interior methods for nonlinear programming", *SIAM J. Optimization* **9** (1999) 84:111, and

S. Wright, "Effects of finite-precision arithmetic on interior-point methods for nonlinear programming", *SIAM J. Optimization* **12** (2001) 36:78

[2] See previous footnote ...

demonstrates that this "defect" is far from fatal.

Interior-point methods

The interior-point revolution was started by

> N. Karmarkar, "A new polynomial-time algorithm for linear programming", *Combinatorica* **4** (1984) 373:395.

It didn't take long for

> P. Gill, W. Murray, M. Saunders, J. Tomlin and M. Wright, "On projected Newton barrier methods for linear programming and an equivalence to Karmarkar's projective method", *Math. Programming*, **36** (1986) 183:209

to realize that this radical "new" approach was actually something that nonlinear programmers had tried (but, most unfortunately, discarded) in the past.

SQP methods

The first SQP method was proposed in the overlooked 1963 Harvard Master's thesis of R. Wilson. The generic linesearch SQP method is that of

> B. Pschenichny, "Algorithms for general problems of mathematical programming", *Kibernetica*, **6** (1970) 120:125,

while there is a much larger variety of trust-region SQP methods, principally because of the constraint incompatibility issue.

Merit functions for SQP

The first use of an exact penalty function to globalize the SQP method was by

> S. Han, "A globally convergent method for nonlinear programming", *J. Optimization Theory and Applics*, **22** (1977) 297:309, and

> M. Powell, "A fast algorithm for nonlinearly constrained optimization calculations", in *Numerical Analysis, Dundee 1977* (G. Watson, ed) Springer Verlag (1978) 144:157.

The fact that such a merit function may prevent full SQP steps was observed N. Maratos in his 1978 U. of London Ph. D. thesis, while methods for combating the Maratos effect were subsequently proposed by

> R. Fletcher, "Second-order corrections for non-differentiable optimization", in *Numerical Analysis, Dundee 1981* (G. Watson, ed) Springer Verlag (1982) 85:114, and

> R. Chamberlain, M. Powell, C. Lemaréchal, and H. Pedersen, "The watchdog technique for forcing convergence in algorithms for constrained optimization", *Math. Programming Studies*, **16** (1982) 1:17.

An SQP method that avoids the need for a merit function altogether by staying feasible is given by

E. Panier and A. Tits, "On combining feasibility, descent and superlinear convergence in inequality constrained optimization",*Mathematical Programming*, **59** (1992) 261;276.

Hessian approximations

There is a vast literature on suitable Hessian approximations for use in SQP methods. Rather than point at individual papers, a good place to start is

P. Boggs and J. Tolle, "Sequential quadratic programming", *Acta Numerica* **4** (1995) 1:51,

but see also our paper

N. Gould and Ph. Toint, "SQP methods for large-scale nonlinear programming", in *System modelling and optimization, methods, theory and applications* (M. Powell and S. Scholtes, eds.) Kluwer (2000) 149:178.

Trust-region SQP methods

Since the trust-region and the linearized constraints may be incompatible, almost all trust-region SQP methods modify the basic SQP method in some way. The Sℓ_1QP method is due to

R. Fletcher, "A model algorithm for composite non-differentiable optimization problems", *Math. Programming Studies*, **17** (1982) 67:76.

Methods that relax the constraints include those proposed by

A. Vardi, "A trust region algorithm for equality constrained minimization: convergence properties and implementation" *SIAM J. Num. Anal.*, **22** (1985) 575:591, and

M. Celis, J. Dennis and R. Tapia, "A trust region strategy for nonlinear equality constrained optimization", in *Numerical Optimization 1984* (P. Boggs, R. Byrd and R. Schnabel, eds), SIAM (1985) 71:82,

as well as a method that appeared in the 1989 U. of Colorado at Boulder Ph. D. thesis of E. Omojokun, supervised by R. Byrd. The highly original Filter-SQP approach was proposed by

R. Fletcher and S. Leyffer, "Nonlinear programming without a penalty function", *Math. Programming A*, **91** (2002) 239:269,

while the analysis of a typical algorithm may be found in

R. Fletcher, S. Leyffer and Ph. Toint, "On the global convergence of a Filter–SQP algorithm", *SIAM J. Optimization* **13** (2002) 44:59.

Modern methods for nonlinear programming

Many modern methods for nonlinearly constrained optimization tend to be SQP-interior-point hybrids. A good example is due to

R. Byrd, J. Gilbert and J. Nocedal, "A trust region method based on interior point techniques for nonlinear programming", *Math. Programming A* **89** (2000) 149:185,

and forms the basis for the excellent KNITRO package.